FRAMEWORK FOR INTEROPERABLE AND DISTRIBUTED EXTRACTION-TRANSFORMATION-LOADING (ETL) BASED ON SERVICE ORIENTED ARCHITECTURE

MOHAMMED M. I. AWAD

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2012

Awang Had Salleh Graduate School of Arts and Sciences
UUM CAS
Engaging Minds for a Better Tomorrow

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

MOHAMMED M. I. AWAD

calon untuk Ijazah PhD
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"FRAMEWORK FOR INTEROPERABLE AND DISTRIBUTED EXTRACTION-TRANSFORMATION-LOADING (ETL) BASED ON SERVICE ORIENTED ARCHITECTURE"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***19 Oktober 2011.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: ***October 19, 2011.***

Pengerusi Viva: *(Chairman for VIVA)*	**Prof. Dr. Norshuhada Shiratuddin**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Assoc. Prof. Dr. Wan Mohd Nasir Wan Kadir**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Dr. Nor Laily Hashim**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Mohd Syazwan Abdullah**	Tandatangan *(Signature)*

Tarikh:
(Date) *October 19, 2011*

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Pengekstrakan, Transformasi dan Proses Pemuatan (ETL) merupakan fungsi-fungsi utama dalam penyelesaian gudang data. Kekurangan komponen pengagihan dan saling kendali menyebabkan wujudnya lompang yang menimbulkan banyak masalah dalam domain ETL. Ini terjadi kerana komponen-komponen dalam kerangka kerja ETL sedia ada adalah saling berkait. Kajian ini membincangkan bagaimana untuk mengagihkan komponen-komponen ETL supaya komponen pengagihan dan saling kendali dapat dilaksanakan. Tambahan pula, kajian ini menunjukkan bagaimana kerangka kerja ETL dapat diperluaskan. Untuk mencapai tujuan tersebut, Perkhidmatan Berorientasikan Seni Bina digunakan untuk memperjelaskan ciri-ciri pengagihan dan saling kendali yang tidak wujud sebelum ini, dengan cara menyusun semula kerangka kerja ETL. Kajian ini menyumbang kepada bidang ETL dengan penambahan konsep pengagihan dan saling kendali kepada kerangka kerja ETL. Seterusnya, kajian ini juga menyumbang kepada bidang penggudangan data dan kepintaran perniagaan kerana ETL merupakan konsep utama dalam bidang ini. Metodologi Design Science Approach (DSA) dan Scrum digunakan untuk mencapai matlamat kajian ini. Integrasi di antara kedua-dua metodologi tersebut dapat mencapai objektif kajian ini. Kerangka kerja ETL yang baru ini direalisasikan menerusi pengujian dan penghasilan satu prototaip yang berdasarkan kepada kerangka kerja tersebut. Kejayaan prototaip ini dinilai berdasarkan tiga kajian kes yang melibatkan data dan alatan daripada tiga organisasi. Organisasi tersebut menggunakan penyelesaian gudang data untuk menghasilkan laporan statistik yang membolehkan pengurusan atasan membuat keputusan. Dapatan ketiga-tiga kajian kes ini menunjukkan komponen pergagihan dan saling kendali dapat dicapai dengan menggunakan kerangka kerja yang baru dalam ETL.

Katakunci: Pengekstrakan, Transformasi dan Proses Pemuatan, Gudang data, Sistem, Perkhidmatan berorientasikan seni bina.

Abstract

Extraction, Transformation and Loading (ETL) are the major functionalities in data warehouse (DW) solutions. Lack of component distribution and interoperability is a gap that leads to many problems in the ETL domain, which is due to tightly-coupled components in the current ETL framework. This research discusses how to distribute the Extraction, Transformation and Loading components so as to achieve distribution and interoperability of these ETL components. In addition, it shows how the ETL framework can be extended. To achieve that, Service Oriented Architecture (SOA) is adopted to address the mentioned missing features of distribution and interoperability by restructuring the current ETL framework. This research contributes towards the field of ETL by adding the distribution and interoperability concepts to the ETL framework. This leads to contributions towards the area of data warehousing and business intelligence, because ETL is a core concept in this area. The Design Science Approach (DSA) and Scrum methodologies were adopted for achieving the research goals. The integration of DSA and Scrum provides the suitable methods for achieving the research objectives. The new ETL framework is realized by developing and testing a prototype that is based on the new ETL framework. This prototype is successfully evaluated using three case studies that are conducted using the data and tools of three different organizations. These organizations use data warehouse solutions for the purpose of generating statistical reports that help their top management to take decisions. Results of the case studies show that distribution and interoperability can be achieved by using the new ETL framework.

Keywords: Extraction, Transformation and Loading, Data warehousing, Service Oriented Architecture.

List of Publications, Invited Speaker and Awards

Awad, M. M. I., & Abdullah, M. S. (2010). A Framework for Interoperable Distributed ETL Components Based on SOA. *Proceeding of the 2nd International Conference on Software Technology and Engineering (ICSTE 2010).*

Awad, M. M. I., & Abdullah, M. S. (2010). Extending ETL Framework using Service Oriented Architecture. *Procedia Computer Science Journal*, (3), 110-114.

Mohammed M I Awad - Keynote Speaker[1]. A Framework for Open-Source Interoperable Distributed ETL Components Based on SOA[2]. *Malaysia Open Source Conference 2010 (MOSC 2010).*

Best Award. *Malaysia Technology Expo 2011 (MTE 2011).*

Gold Medal. *Malaysia Technology Expo 2011 (MTE 2011).*

Silver Medal. *Seoul International Invention Fair 2011 (SIIF 2011),* South Korea.

Gold Medal. *Malaysia Technology Expo 2012 (MTE 2012).*

[1] http://conf.oss.my/speakers.html
[2] http://conf.oss.my/schedule.html

Acknowledgement

By the Name of Allah, the Most Gracious and the Most Merciful

My most profound thankfulness goes to my father and mother who have motivated me since my childhood to help me reach this success point. Thank you my parents and I hope to be the good son who can make you proud of him.

My deep thankfulness goes to my supervisor Dr. Mohd Syazwan Abdullah for supervising me during the journey of this research. I like his way of conducting the research, especially, his instructions for reviewing the literature, and the whole research as well.

Last but not least, I would like to thank my brothers and sisters as well as my friends.

Table of Contents

Permission to Use i
Abstrak ii
Abstract iii
List of Publications, Invited Speaker and Awards iv
Acknowledgement v
Table of Contents vi
List of Tables xi
List of Figures xii
List of Appendices xv
List of Abbreviations xvi

CHAPTER ONE INTRODUCTION 1
1.1 Research Background 1
1.2 Research Motivation 4
1.3 Problem Statement 7
1.4 Research Questions 11
1.5 Research Objectives 11
1.6 Research Strategy 13
1.7 Scope of the Research 15
1.8 Contributions 16
1.9 Thesis Organization 18
1.10 Conclusion 19

CHAPTER TWO EXTRACTION, TRANSFORMATION, AND LOADING IN DATA WAREHOUSING 20
2.1 Data Warehouse 20
2.2 Brief Discussed Overview of traditional ETL Framework Components 22
2.3 Industrial ETL Tools 26
2.4 Component Coupling in Current ETL Structure 30
2.5 Components Distribution 32
2.6 Compatibility of ETL Components with Heterogeneous Environments 33
2.7 Extensibility and Scalability of ETL Framework 36

2.8 ETL Tools Administration 38
2.9 ETL Tools Licensing 40
2.10 Conclusion 41

CHAPTER THREE DISTRIBUTED TECHNOLOGIES 42

3.1 Distributed Systems 42
3.1.1 Remote Procedure Call (RPC) 45
3.1.2 Common Object Request Broker Architecture (CORBA) 46
3.1.3 Distributed Component Object Model (DCOM) 48
3.1.4 Remote Method Invocation (RMI) 49
3.1.5 Service Oriented Architecture (SOA) 51
3.1.6 Comparison of Distributed Technologies 53
3.2 Application Architectures and SOA 57
3.2.1 Types of Application Architectures 58
3.3 Extensible Markup Language (XML) 62
3.4 Web Services in SOA 63
3.4.1 Web Services versus SOA Concepts 63
3.4.2 Simple Object Access Protocol (SOAP) in Web Services 64
3.4.3 Web Services Description Language (WSDL) 64
3.4.4 Messaging 65
3.5 Integration Benefits of Adopting SOA in the ETL Framework 66
3.6 Conclusion 67

CHAPTER FOUR RESEARCH METHODOLOGY 68

4.1 Design Science Approach (DSA) 69
4.1.1 DSA phases 70
4.2 Scrum Methodology 75
4.2.1 Scrum Artifacts 76
4.2.2 SCRUM Phases 77
4.3 New ETL Framework Development Using DSA 78
4.4 Applying Scrum Methodology for the Development Phase of DSA 87
4.5 Conclusion 93

CHAPTER FIVE THE NEW ETL FRAMEWORK 94

5.1 Overview of the New ETL Framework 94
5.2 Distributed Architecture Specifications 100
5.3 Web Services Involvement Specifications 101
5.3.1 A Web Service for every Distributed ETL Component 103
5.3.2 Web Service Operations 103
5.3.3 WSDL Documents 104
5.3.4 XML Schema 107
5.4 Orchestration Point Specifications 109
5.5 Specifications of the Composition of Partners and Configurations Based on SOA 112
5.6 Specifications of Extending the New ETL Framework 115
5.7 Meta-Model for the New ETL Framework 117
5.8 Feedback from the Experts Regarding the Theoretical Framework 128
5.9 Conclusion 131
CHAPTER SIX SOA-BASED ETL PROTOTYPE 133
6.1 Analysis 133
6.1.1 Requirements (Prototype Backlog) Determination 134
6.1.2 Backlog Division 135
6.2 Database Sprint 138
6.3 Coding ETL Components Sprint 143
6.3.1 Extraction Task 145
6.3.2 Transformation Task 147
6.3.3 Classified-Fragmentation Task 149
6.3.4 Loading Task 151
6.4 Distributed Components and Web Services Sprint 153
6.5 Business Process Execution Language (BPEL) Creation Sprint 154
6.6 Sprint of Assembling Prototype Components in One Composite Application 157
6.7 Sprints of Testing 158
6.7.1 Unit Testing Sprint 160
6.7.2 Classified-Fragmentation Speed and Scalability Testing Sprint 161
6.7.3 Compatibility Testing Sprint 163

6.7.4 End To End Testing Sprint 163
6.8 Conclusion 166
CHAPTER SEVEN EVALUATION 167
7.1 Case Study 1: Applying ETL Functionalities on Palestine Electric Company (PEC) using Traditional and New ETL Tools 170
7.1.1 ETL Business Needs 170
7.1.2 Extracting Required Fields for Data Warehouse Star-Schema 171
7.1.3 Applying ETL Functionalities Using the Traditional ETL Tool 172
7.1.4 SOA-based ETL Prototype 174
7.1.5 Goals Achieved 175
7.2 Case Study 2: Applying ETL Functionalities on Limkokwing University of Creative Technology (LUCT) using Traditional and New ETL Tools 179
7.2.1 ETL Business Needs 179
7.2.2 Extracting Required Fields for Data Warehouse Star-Schema 180
7.2.3 Applying ETL Functionalities Using the Traditional ETL Tool 181
7.2.4 SOA-based ETL Prototype 183
7.2.5 Goals Achieved 184
7.3 Case Study 3: Applying ETL Functionalities on Professionals Information Technology (PIT) Company using Traditional and New ETL Tools 188
7.3.1 ETL Business Needs 189
7.3.2 Extracting Required Fields for Data Warehouse Star-Schema 189
7.3.3 Applying ETL Functionalities Using the Traditional ETL Tool 190
7.3.4 SOA-based ETL Prototype 192
7.3.5 Goals Achieved 193
7.4 Conclusion 196
CHAPTER EIGHT CONCLUSION AND FUTURE WORK 197
8.1 Conclusion 197
8.1.1 Problems of Current ETL Framework 198
8.1.2 Proposing the New ETL Framework 199
8.1.3 Defining the New ETL Framework 200
8.1.4 Validating the New ETL Framework 200

List of Tables

Table 2.1: Component Coupling features of Some ETL Tools 31
Table 2.2: Compatibility Issues of Some Popular Commercial ETL Tools 35
Table 2.3: Administration Capabilities of Industrial ETL Tools 39
Table 3.1: Distributed Technologies based on Components Distribution and Interoperability 54
Table 3.2: Distributed Technologies based on Components Portability, Extensibility, and Legacy Compatibility 55
Table 4.1: DSA Methodology Phases 70
Table 4.2: Types of Case Study Evidence (Tellis, 1997) 73
Table 4.3: Research Strategy Based on DSA 79
Table 4.4(Part A): Details of the Case Studies based on Case Study Design 91
Table 4.4 (Cont., Part B): Details of the Case Studies based on Case Study Design 92
Table 5.1: EBNF Symbols Summary (ISO, 1996; Yong Xia, 2002; Gargantini, 2007) 120
Table 5.2: Problems Satisfaction 130
Table 5.3: Solutions Satisfaction 130
Table 5.4: Advantages Satisfaction 131
Table 6.1: Time Deference between Fragmented and Un-Fragmented Data for Report Generation 162
Table 7.1: Summary Report for 2010 Electricity Blackouts 173
Table 7.2: Observation Results Checklist 178
Table 7.3: Summary Report for Students' Performance in Java II for Feb-June, 2011 Semester 182
Table 7.4: Observation Results Checklist 188
Table 7.5: Summary Report for Trainers' Performance in Training IT Courses 191
Table 7.6: Observation Results Checklist 196

List of Figures

Figure 1.1: Research Strategy 13
Figure 2.1: General Data Warehouse Components (Kimball & Caserta, 2004) 21
Figure 2.2: Traditional ETL Framework 23
Figure 3.1: SOA Parts (Barai et al., 2008) 53
Figure 3.2: N-Tier Architecture (Armstrong et al., 2004) 58
Figure 3.3: Data Portability Feature in XML (Barai et al., 2008) 62
Figure 4.1: Flow Chart of the Methodology Applied in this Research 88
Figure 4.2: Case Study Design 90
Figure 5.1: A Conceptual Framework for Interoperable Distributed ETL Components 95
Figure 5.2: Flow Diagram for Steps to Consume an ETL Service by a Client 97
Figure 5.3: ETL Composition Architecture Adopted From (Salter & Jennings, 2008) 114
Figure 5.4: A Framework for Interoperable Distributed ETL Components with Classified - Fragmentation 116
Figure 5.5: Meta-Model for Interoperable and Distributed ETL Framework Components Based on SOA 119
Figure 5.6: GlassFish Server and Containers (Armstrong et al., 2004) 128
Figure 6.1: CLINIC Database 139
Figure 6.2: EXTRACT TEMP STORAGE Database 140
Figure 6.3: TRANSFORM TEMP STORAGE Database 140
Figure 6.4: CLASSIFICATION Database 141
Figure 6.5: LOAD Database 142
Figure 6.6: Class Diagram of ETL Components 144
Figure 6.7: config.txt (JDBC Connection Variables) 145
Figure 6.8: Extraction Sequence Diagram 146
Figure 6.9: Transformation Sequence Diagram 147
Figure 6.10: Classified-Fragmentation Sequence Diagram 149
Figure 6.11: Transformation Sequence Diagram 152
Figure 6.12: Design of the BPEL Orchestration Point 156
Figure 6.13: Design of the Composite Application 158
Figure 6.14: The Main Web Interface of the SOA-based ETL Prototype 159
Figure 6.15: GlassFish Tester Result for the Extract Web Service 160
Figure 6.16: A Statistical Report Generated from a Clinical DW Repository 162

Figure 6.17: End to End Test Case Input File (input.xml) 164
Figure 6.18: Auto generated End to End Test Case Output File (output.xml) 164
Figure 6.19: Sample Data **before** Executing the Transformation Component 165
Figure 6.20: Sample Data **after** Executing the Transformation Component 166
Figure 7.1: GlassFish ESB Based on NetBeans IDE for Managing the SOA-based ETL Prototype 168
Figure 7.2: First Step in Executing Traditional ETL Tools 169
Figure 7.3: Second Step in Executing Traditional ETL Tools 169
Figure 7.4: Third Step in Executing Traditional ETL Tools 169
Figure 7.5: Sample Partial Source PEC DB Schema 171
Figure 7.6: Screenshot of Auto-generated Data in Turbine_transaction_history Table 171
Figure 7.7: Screenshot of Auto-generated Data in Transaction_types Table 171
Figure 7.8: Star Schema of the Fact and Dimension Tables Extracted from PEC Database 172
Figure 7.9: Graph Report for 2010 Electricity Blackouts 174
Figure 7.10: A Sample of the Auto Generated Data of the Fact Table (elec_brkout) of the Star Schema Explored in Figure 7.8 176
Figure 7.11: The Data of the Dimension Table (months) of the Star Schema Explored in Figure 7.8 176
Figure 7.12: The Data of the Dimension Table (periods) of the Star Schema Explored in Figure 7.8 177
Figure 7.13: The Data of the Dimension Table (turbines) of the Star Schema Explored in Figure 7.8 177
Figure 7.14: The Data of the Dimension Table (periods) of the Star Schema Explored in Figure 7.8 (after executing the translation component) 177
Figure 7.15: Sample Partial Source DB Schema of LUCT 180
Figure 7.16: Star Schema of the Fact and Dimension Tables Extracted from LUCT Database 181
Figure 7.17: Summary Report for Students' Performance in Java II for Feb-June, 2011 Semester 183
Figure 7.18: A Sample of the Auto Generated Data of the Fact Table (student_performance) of the Star Schema Explored in Figure 7.16 185
Figure 7.19: The Data of the Dimension Table (marks) of the Star Schema Explored in Figure 7.16 186
Figure 7.20: The Data of the Dimension Table (attendance) of the Star Schema Explored in Figure 7.16 186

Figure 7.21: The Data of the Dimension Table (student_status) of the Star Schema Explored in Figure 7.16 186
Figure 7.22: The Data of the Dimension Table (gender) of the Star Schema Explored in Figure 7.16 186
Figure 7.23: The Data of the Dimension Table (finalexam) of the Star Schema Explored in Figure 7.16 187
Figure 7.24: Sample Partial Source DB schema of PIT 189
Figure 7.25: Star Schema of the Fact and Dimension Tables Extracted from PIT Database 190
Figure 7.26: Graph Report for Trainers' Performance in Training IT Courses. 191
Figure 7.27: A Sample of the Auto Generated Data of the Fact Table (trainer_evaluation) of the Star Schema Explored in Figure 7.25 194
Figure 7.28: The Data of the Dimension Table (trainer_details) of the Star Schema Explored in Figure 7.25 194
Figure 7.29: The Data of the Dimension Table (performance_level) of the Star Schema Explored in Figure 7.25 194

List of Appendices

Appendix A Details of the Prototype Design 220
Appendix B Source Code of the Prototype 275
Appendix C Questionnaires as Deliverables of Structured Interviews Done with Industry Experts 301
Appendix D Case Studies User Manual and Parts of the Source Code 332

List of Abbreviations

Acronym	Description
BI	Business Intelligence
DW	Data Warehouse
ETL	Extraction-Transformation-Loading
SOA	Service Oriented Architecture
J2EE	Java 2 Enterprise Edition
XML	eXtensible Markup Language
SOAP	Simple Object Access Protocol
DSA	Design Science Approach
WSDL	Web Services Description Language
LAN	Local Area Network
DSA	Data Staging Area
OMG	Object Management Group
CORBA	Common Object Request Broker Architecture
RMI	Remote Method Invocation
RPC	Remote Procedure Call
DCOM	Distributed Component Object Model
HTML	Hyper Text Markup Language
JSP	Java Server Pages
EJB	Enterprise Java Beans
DBMS	Database Management System
EIS	Enterprise Information System
HTTP	Hyper Text Transfer Protocol
JMS	Java Messaging Service
OLAP	online analytical processing
IT	Information Technology
OWB	Oracle Warehouse Builder
API	Application Programming Interface

BODI	Business Objects Data Integrator
CAL	client access license
SSIS	SQL Server Integration Services
GUI	Graphical User Interface
BIDS	Business Intelligence Development Studio
DCOM	Distributed Component Object Model
SC	Service Container
JDBC	Java Database Connectivity
BPEL	Business Process Execution Language

CHAPTER ONE
INTRODUCTION

This chapter presents the background and outlines the motivation of the research. This is followed by the research problems, the research question and the research objectives. Moreover, the research strategy is discussed and the scope of the research is argued. Furthermore, the research contribution is highlighted, the organization of the thesis is explored and chapter conclusions are presented.

1.1 Research Background

Data warehouses (DW) have become a main component of the corporate information system architecture, in which it plays a major role in building decision support systems (Vassiliadis *et al.*, 2002; Darmont *et al.*, 2005; Wrembel & Koncilia, 2007). By collecting data from a variety of internal and external sources, data warehouses use the transformation functionality which is a function in the ETL framework (explained in Chapter Two) to provide homogeneous information for analysis and reporting tasks (Wrembel & Koncilia, 2007; Bala *et al.*, 2009).

The uses of data warehousing products and services have been increasing over the years by industry as well as the development of the related technologies (Sen & Sinha, 2007; Wrembel & Koncilia, 2007). Furthermore, within the last decade, data warehouse field has made a vry important step by moving from simple centralized repositories to a platform for data integration and analysis (Vitt *et al.*, 2002; Mundy *et al.*, 2006; Watson & Wixom, 2007; Xi & Hongfeng, 2009). This move is pushing the success of the whole Business Intelligence (BI) field.

The contents of the thesis is for internal user only

REFERENCES

Abrahiem, R. (2007). A New Generation of Middleware Solutions for a Near-Real-Time Data Warehousing Architecture. *Proceedings of Electro/Information Technology International Conference.*

Agrawal, H., Chafle, G., Goyal, S., Mittal, S., & Mukherjea, S. (2008). An Enhanced Extract-Transform-Load System for Migrating Data in Telecom Billing. *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering.*

Agrawal, R., Bayardo, R. J., Gruhl, D., & Papadimitriou, S. (2002). Vinci: A service-oriented architecture for rapid development of web applications. *Computer Networks, 39*(5), 523-539.

Albrecht, A., & Naumann, F. (2008). Managing ETL processes. *Proceedings of New Trends in Information Integration (NTII) Workshop.*

Almeida, J., Almeida, V., Ardagna, D., Francalanci, C., & Trubian, M. (2006). Resource management in the autonomic service-oriented architecture. *Proceedings of IEEE International Conference on Autonomic Computing, 2006. ICAC'06.*

Almeida, M. S., Ishikawa, M., Reinschmidt, J., & Roeber, T. (1999). *Getting started with Data Warehouse and Business Intelligence* (1st ed.). San Jose, California, USA: International Business Machines Corporation.

Apache. (2010). Apache JMeter. Retrieved 22/9/2009, from http://jakarta.apache.org/jmeter/.

Armstrong, E., Ball, J., Bodoff, S., Carson, D. B., Evans, I., Green, D., *et al.* (2004). *The J2EE™ 1.4 Tutorial* (2nd ed.). San Antonio Road Palo Alto, CA, USA: Sun Microsystems.

Atkinson, T. K. H. (2002). Rearchitecting the UML Infrastructure. *ACM Transactions on Modeling and Computer Simulation, 12*(4), 290–321.

Ault, M. (2003). *Oracle Data Warehouse Management: Secrets of Oracle Data Warehousing* (1st ed.). North Carolina, USA: Rampant Techpress.

Badoiu, A., Petrescu, S., Vlad, V., & Botu, A. (2008). Information System for the Management of the Health Services in Romania. *Proceedings of IEEE International Conference for Robotics Automation, Quality and Testing.*

Bala, H., Venkatesh, V., Venkatraman, S., Bates, J., & Brown, S. H. (2009). Disaster Response in Health Care: A Design Extension for Enterprise Data Warehouse. *communications of the acm, 1*(52), 136-140.

Bâra, A., Lungu, I., Velicanu, M., Diaconita, V., & Botha, I. (2008). Improving query performance in virtual data warehouses. *WSEAS Transactions on Information Science and Applications, 5*(5), 632-641.

Barai, M., Binildas, & Caselli, V. (2008). *Service Oriented Architecture with Java* (1st ed.). Birmingham, UK: Packt Publishing.

Barton, B., & Campbell, E. (2007). Implementing a Professional Services Organization Using Type C Scrum. *Proceedings of 40th Hawaii International Conference on System Sciences.*

Bertrand, F., Bramley, R., Sussman, A., Bernholdt, D. E., Kohl, J. A., Larson, J. W., *et al.* (2005). Data redistribution and remote method invocation in parallel component architectures. *Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium.*

Blair, G. S., Coulson, G., Robin, P., & Papathomas, M. (2009). An architecture for next generation middleware. *Proceedings of the IFIP International Conference on Distributed Systems Platforms and Open Distributed Processing*, 191-206.

Bonifati, A., Casati, F., Dayal, U., & Shan, M. C. (2001). Warehousing workflow data: Challenges and opportunities. *Proceedings of the International Conference on Very Lagr Databases*, 649-652.

Brace, I. (2008). *Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research* (2nd ed.). London, UK: Kogan Page.

Brown, A., Johnston, S., & Kelly, K. (2003). Using service-oriented architecture and component-based development to build web service applications. *interactions, 1*(1), 2-16.

Bruckner, R. M., List, B., & Schiefer, J. (2002). Striving towards near real-time data integration for data warehouses. *Lecture notes in computer science, 1*(1), 317-326.

Bugatti, P. H., Ribeiro, M. X., Traina, A. J. M., & Jr, C. T. (2008). Content-based Retrieval of Medical Images by Continuous Feature Selection. *Proceedings of 21st IEEE International Symposium on Computer-Based Medical Systems.*

Channabasavaiah, K., Holley, K., & Tuggle, E. (2003). Migrating to a service-oriented architecture. *IBM DeveloperWorks, 1*(1), 1-23.

Cheng, H. K., Tang, Q. C., & Zhao, a. J. L. (2006). Web Services and Service-Oriented Application Provisioning: An Analytical Study of Application Service Strategies. *Proceedings of IEEE International Conference on Transactions on Engineering Management.*

Chester, T. M. (2001). Cross-Platform Integration with XML and SOAP. *IT Pro Journal, 3*(5), 26 - 34.

Cleveland, F. M. (2002). Information Exchange Modeling (IEM) and extensible Markup Language (XML) Technologies. *Proceedings of IEEE Power Engineering Society Winter Meeting, 2002.*

Coulouris, G., Dallimore, J., & Kindberg, T. (2001). *Distributed systems: concepts and design* (3rd ed.). Shanghai,China: China Machine Press.

Cristal, M., Wildt, D., & Prikladnicki, R. (2008). Usage of SCRUM Practices within a Global Company. *Proceedings of IEEE International Conference on Global Software Engineering.*

Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N., & Weerawarana, S. (2002). Unraveling the Web services web: an introduction to SOAP, WSDL, and UDDI. *IEEE Internet Computing, 6*(2), 86-93.

Darmont, J., & Boussaid, O. (2006). *Processing and managing complex data for decision support* (1st ed.). Hershey, USA: IGI Global.

Darmont, J., Boussaid, O., Ralaivao, J. C., & Aouiche, K. (2005). An architecture framework for complex data warehouses. *Proceedings of 7th International Conference on Enterprise Information Systems.*

Derong, S., Ge, Y., Yu, C., Yue, K., & Tiezheng, N. (2005). An Effective Web Services Discovery Strategy for Web Services Composition.

Proceedings of the 2005 The Fifth International Conference on Computer and Information Technology (CIT'05).

Dessloch, S., Hemaindezt, M. A., Wisneskyl, R., Radwan, A., & Zhou, J. (2008). Orchid: Integrating Schema Mapping and ETL. *Proceedings of IEEE Data Engineering 24th International Conference.*

Dou, A. J., Lin, S., & Kalogeraki, V. (2008). Real-Time Querying of Historical Data in Flash-equipped Sensor Devices. *Real-Time Systems Symposium.*

Du, D., & Raghavendra, C. (2005). *Distributed Network Systems* (2nd ed.). Califorina, USA: Springer.

Du, T. C., & Wong, J. (2004). Designing Data Warehouses for Supply Chain Management. *Proceedings of IEEE International Conference on E-Commerce Technology.*

Dung, T. Q., & Kameyama, W. (2007). A Proposal of Ontology-based Health Care Information Extraction System: VnHIES. *Proceedings of Research, Innovation and Vision for the Future, 2007 IEEE International Conference.*

Erl, T. (2004). *Service-oriented architecture: a field guide to integrating XML and web services* (1st ed.). NJ, USA: Prentice Hall.

Gao, A., Yang, D., & Tang, S. Web service composition based on message schema analysis. *Advances in Databases: Concepts, Systems and Applications, 4443*(5), 918-923.

Gargantini, E. R., and Patrizia Scandurra. (2007). Deriving a textual notation from a metamodel by University of Bergamo. Retrieved 1/12/2011, from https://doc.telin.nl/dsweb/Get/Rendition-50041/3M4MDA_2006_online_proceedings.pdf#page=41.

Green, P. F., Indulska, M. K., Rosemann, M., & Weber, R. A. (2003). Will XML technologies and web services solve the interoperability problem? *Proceedings of International Workshop on Utility, Usability and Complexity of Emergent IS.*, 103-115.

Hau, T., Ebert, N., Hochstein, A., & Brenner, W. (2008). Where to Start with SOA Criteria for Selecting SOA Projects. *Proceedings of 41st Hawaii International Conference on System Sciences.*

He, B., Wang, R., Chen, Y., Lelescu, A., & Rhodes, J. (2007). BIwTL: a business information warehouse toolkit and language for warehousing simplification and automation. *Proceedings of the 2007 ACM SIGMOD international conference on Management of data.*

Heinzl, S., Mathes, M., Friese, T., Smith, M., & Freisleben, B. (2006). Flex-SwA: Flexible Exchange of Binary Data Based on SOAP Messages with Attachments. *Proceedings of IEEE International Conference on Web Services (ICWS'06).*

Henry, S., Hoon, S., Hwang, M., Lee, D., & DeVore, M. D. (2005). Engineering Trade Study: Extract, Transform Load Tools for Data Migration. *Proceedings of the 2005 Systems and Information Engineering Design Symposium.*

Hevner, A. R., & Chatterjee, S. (2010). *Design Research in Information Systems* (1st ed.). CA, USA: Springer.

Hevner, A. R., & March, S. T. (2003). The Information Systems Research Cycle. *MIS Quarterly (November), 36*(11), 111 - 113.

Holzer, S., Tafazzoli, A. G., Altmann, U., Wachter, W., & Dudeck, J. (1999). *Data warehousing as a tool for quality management in oncology* (1st ed.). NY, USA: IOS Press.

IBM. (2010). InfoSphere DataStage. Retrieved 15/2/2010, from http://www-01.ibm.com/software/data/infosphere/datastage/.

Information, C. I. f. H. (2009). Regrouping Historical Data CIHI Reference Document. *Canadian Institute for Health Information* Retrieved 12/2/2010, from http://www.cihi.ca/CIHI-ext-portal/pdf/internet/REGROUPING_HIST_DATA_REF_EN.

Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). Indiana, USA: Wiley Publishing, Inc.

Iqbal, T., & Daudpota, N. (2006). XML based framework for ETL processes for relational databases. *WSEAS Transactions on Information Science and Applications, 3*(7), 1402-1406.

ISO. (1996). *Extended Backus-Naur Form (EBNF)* Retrieved 30/10/2011, from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26153.

Issarny, V., Kloukinas, C., Zarras, A., & Architectures, M. (2008). Management Group's Common Object Request Broker (CORBA). *Microsoft's Distributed Component Object Model.*

Jerstad, I., Dustdar, S., & Thanh, D. V. (2005). A service oriented architecture framework for collaborative services. *Proceedings of 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise, 2005.*

Johansen, M. A., Scholl, J., Hasvold, P., Ellingsen, G., & Bellika, J. G. (2008). "Garbage In, Garbage Out"- Extracting Disease Surveillance Data from EPR Systems in Primary Care. *Proceedings of 2008 ACM conference on Computer supported cooperative work.*

Judy, K. H., & Krumins-Beens, I. (2008). Great Scrums Need Great Product Owners: Unbounded Collaboration and Collective Product Ownership. *Proceedings of 41st Hawaii International Conference on System Sciences.*

Katifori, A., Torou, E., Vassilakis, C., & Halatsis, C. (2008). Supporting Research in Historical Archives:Historical Information Visualization Modeling Requirements. *Proceedings of 12th International Conference Information Visualisation.*

Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit* (2nd ed.). Indiana, USA: Wiley Publishing, Inc.

Klmek, J., Kopenec, L., Loupal, P., & Mal, J. (2010). XCase - A Tool for Conceptual XML Data Modeling. *Proceedings of Advances in Databases and Information Systems Conference.*

Kobryn, C. (2000). Modeling Components and Frameworks with UML. *Communications of the ACM, 43*(10).

Kruchten, W. K., Bran, & Selic. (2001). Describing Software Architecture with UML. *Rational Software.*

Kshemkalyani, A. D., & Singhal, M. (2008). *Distributed Computing Principles, Algorithms, and Systems* (1st ed.). Cambridge, UK: Cambridge University Press.

Kumari, G. P., Kandan, B., & Mishra, A. K. (2008). Experience sharing on SOA based Heterogeneous Systems Integration. *IEEE Congress on Services 2008 - Part I, 11*(6), 107 - 108.

Lam, T., & Minsky, N. (2010). Regulating Orchestration in SOA-Based Systems. *Proceedings of 2010 Seventh International Conference on Information Technology.*, 690-695.

Laskey, K. B., & Laskey, K. (2009). *Service oriented architecture*. CA, USA: Computational Statistics.

Li, G., Muthusamy, V., & Jacobsen, H. A. (2010). A distributed service-oriented architecture for business process execution. *ACM Transactions on the Web (TWEB), 4*(1), 1-33.

Liao, H., Padmanabhan, S., Srinivasan, S., Lau, P., Shan, J., & Wisnesky, R. (2008). Bringing Business Objects into Extract-Transform-Load (ETL) Technology. *Proceedings of IEEE International Conference on e-Business Engineering*.

Louridas, P. (2006). SOAP and Web Services. *IEEE Software*.

Lujanmora, J. T. (2004). Physical Modeling of Data Warehouses using UML. *ACM Journal*.

Maassen, J., Nieuwpoort, R., Veldema, R., Bal, H. E., & Plaat, A. (2008). *Java Remote Method Invocation provides an unusually flexiblilty*. Indiana, USA: Wiley Publishing, Inc.

Mahboubi, H., & Darmont, J. e. o. (2009). Enhancing XML Data Warehouse Query Performance by Fragmentation. *2009 ACM symposium on Applied Computing*.

Marchenko, A., & Abrahamsson, P. (2008). Scrum in a Multiproject Environment:An Ethnographically-Inspired Case Study on the Adoption Challenges. *Proceedings of Agile 2008 Conference*.

Maskat, R., & Shamsudin, M. F. (2008). Proposing a Physical Model for Malaysian Health Plan Data. *Proceedings of International Conference on Advanced Computer Theory and Engineering*.

Massachusetts, E. D. (2008). *Introduction to the Data Warehouse* (1st ed.). Malden, USA: Massachusetts.

Matsumura, I., Ishida, T., Murakami, Y., & Fujishiro, Y. (2006). Situated Web Service: Context-Aware Approach to High-Speed Web Service Communication. *Proceedings of IEEE International Conference on Web Services (ICWS'06)*.

Maurizio, A., Sager, J., Jones, P., Corbitt, G., & Girolami, L. (2008). Service Oriented Architecture: Challenges for Business and Academia. *Proceedings of 41st Hawaii International Conference on System Sciences*.

McCabe, M. C., & Grossman, D. (1996). The role of tools in development of a data warehouse. *Proceedings of the 4th International Symposium on Assessment of Software Tools*.

Microsoft. (2009). Microsoft Business Intellegence. Retrieved 5/8/2009, from http://www.microsoft.com/bi/.

Microsoft. (2010). SQL Server Integration Services. Retrieved 15/1/2010, from http://msdn.microsoft.com/en-us/library/ms141026.aspx.

Mitchell, S., Blake, M. B., Cunningham, D., & Gopalan, S. (2008). A SOA-Driven Content Discovery and Retrieval Platform. *Proceedings of 10th IEEE Conference on E-Commerce Technology*.

Morris, H., Liao, H., Padmanabhan, S., Srinivasan, S., Lau, P., Shan, J., *et al.* (2008). Bringing Business Objects into Extract-Transform-Load (ETL) Technology. *Proceedings of the 2008 IEEE International Conference on e-Business Engineering*.

Mrunalini, M., Kumar, T. V. S., & Kanth, K. R. (2009). Simulating Secure Data Extraction in Extraction Transformation Loading (ETL) Processes. *Proceedings of the 2009 Third UKSim European Symposium on Computer Modeling and Simulation*.

Mulik, S., Ajgaonkar, S., & Sharma, K. (2008). Where Do You Want to Go in Your SOA Adoption Journey? *IEEE Computer Society Washington, DC, USA*.

Mundy, J., Thornthwaite, W., & Kimball, R. (2006). *The Microsoft data warehouse toolkit: with SQL Server 2005 and the Microsoft Business Intelligence toolset*: Wiley Pub.

Muñoz, L., Mazón, J.-N., & Trujillo, J. (2009). Measures for ETL processes models in data warehouses. *Proceedings of the first international workshop on Model driven service engineering and data quality and security*.

Mykknen, J., Porrasmaa, J., Rannanheimo, J., & Korpela, M. (2003). A process for specifying integration for multi-tier applications in healthcare. *International journal of medical informatics, 70*(2-3), 173-182.

Natis, Y. (2003). Service-oriented architecture scenario. *Gartner, Inc., Stamford*.

Nelson, G., & Wright, J. (2005). Real time decision support: creating a flexible architecture for real time analytics. *DSSResources. COM, 11*(4), 18-32.

NetBeans. (2010). NetBeans IDE. Retrieved 25/4/2010, from http://netbeans.org/.

Newcomer, E. (2002). *Understanding Web Services: XML, Wsdl, Soap, and UDDI*: Addison-Wesley Professional.

Newcomer, E., & Lomow, G. (2004). *Understanding SOA with Web Services (Independent Technology Guides)* (1st ed.). USA: Addison-Wesley Professional.

Niehaves, B., & Becker, J. (2006). Design Science Perspective on IT-Consulting. *Tagungsband / der Multikonferenz Wirtchaftsinformatik*.

OMG. (2003). *Common Warehouse Metamodel (CWM) Specification* (1st ed. Vol. 1). MA, USA: OMG Headquarters.

OMG. (2011). Unified Modeling Language. Retrieved 29/10/2011, from http://www.uml.org/.

Oracle. (2009). Oracle Enterprise Performance Management and Business Intelligence. Retrieved 10/8/2009, from http://www.oracle.com/solutions/business_intelligence/index.html.

Oracle. (2010). Oracle Warehouse Builder. Retrieved 3/2/2010, from http://www.oracle.com/technetwork/developer-tools/warehouse/overview/index.html.

Papazoglou, M. P. (2003). Service-oriented computing: Concepts, characteristics and directions. *Proceedings of the Fourth International Conference on Web Information Systems Engineering.*

Patasiene, I., Kregzdyte, R., Patasius, M., Patasius, J., & Kazakeviciute, A. (2007). Integrating Global Data into Local Health Data Base. *Proceedings of 29th Annual International Conference of the IEEE EMBS Cité Internationale.*

Pentaho. (2006). Pentaho Data Integration: Spoon 2.3.1, User Manual. Retrieved 3/1/2009 from http://www.pentaho.org.

Pentaho. (2009). Pentaho Business Intellegence. Retrieved 2/7/2009, from http://www.pentaho.com.

Perin, F. (2009). Enabling the Evolution of J2EE Applications through Reverse Engineering and Quality Assurance. *Proceedings of the 2009 16th Working Conference on Reverse Engineering.*

Phan, T., Han, J., Schneider, J.-G., Ebringer, T., & Rogers, T. (2008). A survey of policy-based management approaches for Service Oriented Systems. *Proceedings of 19th Australian Conference on Software Engineering.*

Priebe, T., & Pernul, G. (2003). Towards integrative enterprise knowledge portals. *Proceedings of the twelfth international conference on Information and knowledge management.*

Qiu, B., Liu, Y., Ong, Y. S., Gooi, H. B., & Chen, S. (2002). Managing Metadata over the WWW using eXtensible Markup Language (XML). *Power Engineering Society Winter Meeting, 2002.*

Rayhan, S. H., & Haque, N. (2008). Incremental Adoption of Scrum for Successful Delivery of an IT Project in a Remote Setup. *Proceedings of Agile 2008 Conference.*

Roach, T., Low, G., & D'Ambra, J. (2008). CAPSICUM - A Conceptual Model for Service Oriented Architecture. *IEEE Congress on Services 2008 - Part I.*

Roy, N., Pallapa, G., & Das, S. K. (2008). An Ontology-Driven Ambiguous Contexts Mediation Framework for Smart Healthcare Applications. *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments.*

Sahama, T. R., & Croll, P. R. (2007). A Data Warehouse Architecture for Clinical Data Warehousing. *Proceedings of 2007 Conference of Research and Practice in Information Technology*.

Salter, D., & Jennings, F. (2008). *Building SOA-Based Composite Applications Using NetBeans IDE 6* (1st ed.). USA: PACKT Publishing.

Santos, R. J., & Bernardino, J. (2008). Real-Time Data Warehouse Loading Methodology. *Proceedings of the 2008 international symposium on Database engineering & applications*.

SAS. (2010). SAS Enterprise Data Integration Server. Retrieved 15/1/2010, from http://www.sas.com/technologies/dw/entdiserver/index.html.

Schwaber, K. (2009, 2/3/2009). Scrum Development Process. from http://jeffsutherland.com/oopsla/schwapub.pdf.

Sellis, T. (2006). Formal specification and optimization of ETL scenarios. *Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*.

Sen, A., & Sinha, A. P. (2007). Toward Developing Data Warehousing Process Standards: An Ontology-Based Review of Existing Methodologies. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions 37*(1), 17 - 31.

Shah, R. C., Roy, S., Jain, S., & Brunette, W. (2003). Data mules: Modeling and analysis of a three-tier architecture for sparse sensor networks. *Ad Hoc Networks, 1*(2-3), 215-233.

Shaikh, M. U., Malik, S. U. R., Qureshi, A., & Yaqoob, S. (2010). Intelligent Decision Making Based on Data Mining Using Differential Evolution Algorithms and Framework for ETL Workflow Management. *Proceedings of the 2010 Second International Conference on Computer Engineering and Applications - Volume 01*.

Shani, U., Sela, A., Akilov, A., Skarbovski, I., & Berk, D. (2006). A scalable heterogeneous solution for massive data collection and database loading. *Proceedings of the 1st international conference on Business intelligence for the real-time enterprises*.

Shil, A. B., & Ahmed, M. B. (2006). Additional Functionalities to SOAP, WSDL and UDDI for a Better Web Services' Administration. *Proceedings of Information and Communication Technologies Conference*.

Silvers, F. (2008). *Building and Maintaining a Data Warehouse* (1st ed.). Philadelphia, USA: Taylor & Francis Group, LLC.

Simitsis, A., Skoutas, D., & Castellanos, M. (2010). Representation of conceptual ETL designs in natural language using Semantic Web technology. *Data & Knowledge Engineering: Elsevier Science Publishers, 69*(1), 96 -15.

Simitsis, A., Vassiliadis, P., Terrovitis, M., & Skiadopoulos, S. (2005). Graph-based modeling of ETL activities with multi-level transformations and updates. *Lecture notes in computer science, 3589*(1), 43-61.

Simitsis, A., Wilkinson, K., Castellanos, M., & Dayal, U. (2009). QoX-driven ETL design: reducing the cost of ETL consulting engagements. *SIGMOD '09: Proceedings of the 35th SIGMOD international conference on Management of data.*

Siqueira, T. L. L., Ciferri, R. R., Times, V. C., & Ciferri, C. D. d. A. (2009). A Spatial Bitmap-based Index for Geographical Data Warehouses. *Proceedings of the 2009 ACM symposium on Applied Computing.*

Skoutas, D., Simitsis, A., & Sellis, T. (2009). Ontology-driven conceptual design of ETL processes using graph transformations. *Journal on Data Semantics XIII, 5530*(1), 120-146.

Sneed, H. M. (2006). Integrating legacy Software into a Service oriented Architecture. *Proceedings of the 10th European Conference on Software Maintenance and Reengineering.*

Sprott, D., & Wilkes, L. (2004). Understanding service-oriented architecture. Retrieved 25/3/2011, from http://msdn.microsoft.com/en-us/library/aa480021.aspx.

Stal, M. (2006). Using architectural patterns and blueprints for service-oriented architecture. *IEEE software, 23*(2), 54-61.

Stojanovic, Z., Dahanayake, A., & Sol, H. (2004). Modeling and design of service-oriented architecture. *Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics.*

Stonebraker, M., & Hellerstein, J. M. (2001). Content integration for e-business. *ACM SIGMOD Record, 30*(2), 552-560.

Sulaiman, T., Barton, B., & Blackburn, T. (2006). Proceedings of AgileEVM - Earned Value Management in Scrum Projects. *AGILE 2006 Conference.*

Sullins, B. G., & Whipple, M. B. (2005). *EJB Cookbook* (1st ed.). Philadelphia, USA: Manning.

Sun-Microsystems. (2010). Glassfish. Retrieved 25/4/2010, from https://glassfish.dev.java.net/.

Sutherland, J., Schoonheim, G., & Rijk, M. (2009). Fully Distributed Scrum: Replicating Local Productivity and Quality with Offshore Teams. *Proceedings of 42nd Hawaii International Conference on System Sciences.*

Sutherland, J., Schoonheim, G., Rustenburg, E., & Rijk, M. (2008). Fully Distributed Scrum: The Secret Sauce for Hyperproductive Offshored Development Teams. *Proceedings of Agile 2008 Conference.*

Sutherland, J., Viktorov, A., Blount, J., & Puntikov, N. (2007). Distributed Scrum: Agile Project Management with Outsourced Development Teams. *Proceedings of 40th Hawaii International Conference on System Sciences.*

Suzumura, T., Yasue, T., & Onodera, T. (2010). Scalable performance of system S for extract-transform-load processing. *Proceedings of the 3rd Annual Haifa Experimental Systems Conference.*

Tam, L. N. (2010). *IBM Data Warehouse and Business Intelligence Solutions* (1st ed.). Vietnam: IBM Software Group.

Tanenbaum, A. S., & Van Steen, M. (2002). *Distributed systems* (2nd ed.). New Jersey, USA: CiteSeer.

Tellis, W. (1997). *Application of a Case Study Methodology* (3rd ed.). CA: Sage Publishing.

Temenos. (2005). *Data Warehouse* (2nd ed.). NY, USA: Temenos USA Inc.

Theodosi, A. D., & Tsihrintzis, G. A. (2008). Using Agents for Feature Extraction: Content Based Image Retrieval for Medical Applications. *Manuscript.*

Thomsen, C., & Pedersen, T. B. (2009). A powerful programming framework for extract-transform-load programmers. *Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP.*

Trujillo, J., & Lujnmora, S. (2003). A UML based approach for modeling ETL processes in data warehouses. *Lecture Notes in Computer Science, 22*(5), 307-320.

Tsai, W. T., Paul, R., Wang, Y., Fan, C., & Wang, D. (2002). Extending WSDL to Facilitate Web Services Testing. *Proceedings of the 7th IEEE International Symposium on High Assurance Systems Engineering (HASE'02).*

Tsenov, M. (2007). Example of communication between distributed network systems using web services. *Proceedings of the 2007 international conference on Computer systems and technologies.*

Tziovara, V., Vassiliadis, P., & Simitsis, A. (2007). Deciding the Physical Implementation of ETL Workflows. *Proceedings of ACM tenth international workshop on Data warehousing and OLAP.*

Urgaonkar, B., Pacifici, G., Shenoy, P., Spreitzer, M., & Tantawi, A. (2005). *An analytical model for multi-tier internet services and its applications.*

Uy, E., & Ioannou, N. (2008). Growing and Sustaining an Offshore Scrum Engagement. *Proceedings of Agile 2008 Conference.*

Vaishnavi, V., & Kuechler, B. (2004). Design Research in Information Systems. from http://home.aisnet.org/displaycommon.cfm?an=1&subarticlenbr=279#designResearchMethodology.

Vara, J. M., Castro, V. d., & Marcos, E. (2005). WSDL automatic generation from UML models in a MDA framework. *Proceedings of the International Conference on Next Generation Web Services Practices (NWeSP '05)*.

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual modeling for ETL processes. *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*.

Vassiliadis, P., Simitsis, A., Terrovitis, M., & Skiadopoulos, S. (2005). Blueprints and measures for ETL workflows. *Lecture notes in computer science, 28*(3), 380-385.

Vitt, E., Luckevich, M., & Misner, S. (2002). *Business intelligence*: Microsoft Press.

Voth, G. R., Kindel, C., & Fujioka, J. (1998). Distributed application development for three-tier architectures: Microsoft on Windows DNA. *IEEE Internet Computing, 2*(2), 41-45.

W3C. (1999). XML Schema Requirements. Retrieved 10/12/2009, from http://www.w3.org/TR/NOTE-xml-schema-req.

W3C. (2010). The amazing em unit and other best practices. Retrieved 1-1-2010, from http://www.w3.org/WAI/GL/css2em.htm.

Wang, C., & Liu, S. (2008). SOA Based Electric Power Real-time Data Warehouse. *Proceedings of Workshop on Power Electronics and Intelligent Transportation System*.

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *COMPUTER-IEEE COMPUTER SOCIETY-, 40*(9), 96.

Weerawarana, S., Curbera, F., Leymann, F., Storey, T., & Ferguson, D. F. (2005). *Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging and More* (1st ed.). NY, USA: Prentice Hall.

Wehrle, P., Miquel, M., & Tchounikine, A. (2005). A Model for Distributing and Querying a Data Warehouse on a Computing Grid. *Proceedings of 2005 11th International Conference on Parallel and Distributed Systems*.

Wehrle, P., Miquel, M., & Tchounikine, A. (2007). A Grid Services-Oriented Architecture for Efficient Operation of Distributed Data Warehouses on Globus. *Proceedings of 21st International Conference on Advanced Networking and Applications*.

Werner, C., Buschmann, C., & Fischer, S. (2004). Compressing SOAP Messages by using Differential Encoding. *Proceedings of the IEEE International Conference on Web Services (ICWS'04)*.

Wolter, R. (2001). XML Web services basics. *Microsoft Developer Network, 12*(1), 66-86.

Wrembel, R., & Koncilia, C. (2007). *Data Warehouses and OLAP: Concepts, Architectures and Solutions* (1st ed.). NY, USA: IRM Press.

Wu, L., Barash, G., & Bartolini, C. (2007). A Service-oriented Architecture for Business Intelligence. *Proceedings of IEEE International Conference on Service-Oriented Computing and Applications (SOCA'07).*

Xi, X., & Hongfeng, X. (2009). Developing a Framework for Business Intelligence Systems Integration Based on Ontology. *International Conference on Networking and Digital Society.*

Yang, C. L., Chang, Y. K., & Chu, C. P. (2008). A Gateway Design for Message Passing on The SOA Healthcare Platform. *IEEE International Symposium on Service-Oriented System Engineering.*

Yin, R. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage Publishing.

Yingying, X., Hao, T., & Peiren, Z. (2003). An Advanced Text-To-Speech Server System Based on SOAP Protocol. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03).*

Yong Xia, M. G. (2002). Rigorous EBNF-based Definition for a Graphic Modeling Language. *Winterthurerstr. 190, CH-8057 Zurich, Switzerland.*

Zhang, & Gracanin, D. (2008). Service-Oriented-Architecture based Framework for Multi-User Virtual Environments. *Proceedings of 2008 Winter Simulation Conference.*

Zhang, & Wang, S. (2008). A Framework Model Study for Ontology-driven ETL Processes. *Proceedings of 4th International Conference on Wireless Communications, Networking and Mobile Computing.*

Zhang, L. J., Zhou, N., Chee, Y. M., Jalaldeen, A., Ponnalagu, K., Sindhgatta, R. R., *et al.* (2008). SOMA-ME: A platform for the model-driven design of SOA solutions. *IBM Systems Journal, 47*(3), 397-413.

Zhang, Q., Li, K., & Yu, J.-h. (2006). Application of Multi-Agent System On Web-Based Data Warehouse for Pricing System of Power Supplier. *Proceedings of Power Systems Conference and Exposition, 2006. PSCE '06.*

Zhou, X., Liu, B., Wang, Y., Zhang, R., Li, P., Chen, S., *et al.* (2008). Building Clinical Data Warehouse for Traditional Chinese Medicine Knowledge Discovery. *Proceedings of International Conference on BioMedical Engineering and Informatics.*

Zhu, Y., An, L., & Liu, S. (2008a). Data Updating and Query in Real-time Data Warehouse System. *Proceedings of International Conference on Computer Science and Software Engineering.*

Zhu, Y. Q., Min, B., & Wei, H. (2008b). The Research of Methodology in Models Mapping for ETL Processes Based on Model Driven. *Information Science and Engieering, 2008. ISISE '08. International Symposium*.